STATISTIQUES

1) EXEMPLES

Suite à un contrôle de mathématiques , dans une classe de 28 élèves , le professeur s'est amusé à classer certains résultats dans trois tableaux :

TABLEAU 1

notes	4	5	6	7	8	9	10	11	12	13	14	15	16	17
nb d'élèves (effectifs)	2	0	2	1	2	4	3	2	4	3	2	0	2	1
effectifs cumulés (croissants)														
effectifs cumulés (décroissants)														
fréquences														
fréquences cumulées (croissantes)														
fréquences cumulées (décroissantes)														

TABLEAU 2

soin des copies	très propre	propre	brouillon
effectifs	7	11	10

TABLEAU 3

temps t de révision en h	[0 ; 1 [	[1 ; 2 [	[2 ; 3 [	[3 ; 4 [	[4 ; 5 [
Centre des classes					
effectifs	7	5	5	6	5
effectifs cumulés (croissants)					
fréquences en %					
fréquences cumulées(croissantes) en					

2) GENERALITES

A) SERIE STATISTIQUE A UNE VARIABLE

Etudier une série statistique à une variable revient à étudier un aspect particulier des éléments d'un ensemble donné.
L'ensemble étudié s'appelle **la population** et chaque élément de l'ensemble s'appelle **un individu**.
Si l'ensemble étudié est trop vaste, on en restreint l'étude à une partie appelée **échantillon**.
L'aspect étudié s'appelle **la variable** ou **le caractère**.

Ex :
La population est la classe ; un individu est un élève de la classe ; la variable étudiée est soit les notes (tab1) , soit le soin des copies (tab2) , soit le temps de révision (tab3) . On détermine ainsi trois séries statistiques à une variable.

B) VARIABLE OU CARACTERE

Une variable (ou caractère) peut prendre plusieurs **valeurs** (ou **modalités**) (ex : Une note peut valoir 15 (tab1) , une copie peut être propre (tab2))
On distingue deux types de caractères :

- Un caractère est **qualitatif** si les modalités ne sont pas des nombres. (ex : Le soin des copies (tab2))

Rem: On attribue parfois des valeurs numériques aux modalités. (ex: pour les départements , 57 désigne la Moselle)

- Un caractère est **quantitatif** si les modalités sont des nombres.
  - Les notes représentent un caractère **quantitatif discrèt**.
    (Les valeurs prises sont isolées 4 , 5 , … , 16 , 17)
  - Le temps de révision est un caractère **quantitatif continu**.
    (Le temps de révision pourrait être n'importe quel nombre t, tel que $2 \leq t < 3$ par exemple.
    Les valeurs de ce caractère sont regroupées en **classes** ([0 ; 1 [; [1 ; 2 [...)
    L'amplitude des classes n'est pas forcément la même . En général, on fait l'hypothèse d'une répartition uniforme à l'intérieur de chaque classe ...

C) EFFECTIFS, FREQUENCES

Le nombre d'individus pour lesquels la variable (le caractère) prend une valeur (modalité) donnée s'appelle **l'effectif** de cette valeur.
Le nombre d'individus de la population est **l'effectif total**.

La fréquence d'une valeur est le rapport entre l'effectif de cette valeur et l'effectif total.
(Une fréquence est souvent donnée en pourcentage ; une fréquence est un nombre compris entre 0 et 1; la somme de toutes les fréquences est 1 ou 100%)

Dans le cas d'une variable quantitative, on peut ordonner les différentes valeurs de la plus petite à la plus grande (ou de la plus grande à la plus petite) puis additionner les effectifs successifs : on obtient ainsi **les effectifs cumulés croissants** (ou décroissants).
On obtient de la même façon **les fréquences cumulées croissantes** (ou décroissantes).

3) Caracteristiques de position : MODE, MOYENNE et MEDIANE

Ce sont des valeurs « centrales » autour desquelles se répartissent les valeurs du caractère.

Notations :

- $x_1, x_2, \ldots, x_p$ sont les valeurs ou les centres des classes si ces valeurs sont regroupées en classe.
- $n_1, n_2, \ldots, n_p$ sont les effectifs respectifs des valeurs $x_1, x_2, \ldots, x_p$
- $f_1, f_2, \ldots, f_p$ sont les fréquences respectives des valeurs $x_1, x_2, \ldots, x_p$
- n est l'effectif total : $n = n_1 + n_2 + \ldots + n_p$

A) MODE (ou CLASSE MODALE)

Le mode , pour un caractère discrèt, est la valeur du caractère, notée Mo, qui correspond au plus grand effectif.
Pour un caractère continu, on parle de **classe modale**. Si les classes ont la même amplitude la classe modale est la classe qui correspond au plus fort effectif.
Il peut y avoir plusieurs modes (ou plusieurs classes modales)

Ex : (tab1) Modes :

Ex : (tab3) Classe modale :

B) MOYENNE

La moyenne des n nombres $x_1, x_2, \ldots, x_n$ est le nombre réel, noté $\overline{x}$, tel que :

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

Les nombres x_i ne sont pas en général distincts deux à deux . En regroupant les x_i égaux entre eux, on obtient :

$$\overline{x} = \frac{n_1 x_1 + n_2 x_2 + \ldots + n_p x_p}{n} = \frac{1}{n}\sum_{i=1}^{p} n_i x_i \quad \text{ou} \quad \overline{x} = f_1 x_1 + f_2 x_2 + \ldots + f_p x_p = \sum_{i=1}^{p} f_i x_i$$

Ex : (tab1) $\overline{x} =$

Dans le cas d'un caractère quantitatif continu la moyenne est calculée en choisissant comme valeur du caractère le centre de la classe.

Ex : (tab3) $\overline{x} =$

Rem :
Si dans une série de notes, une note apparait de manière exceptionnelle (0 par exemple) , on peut calculer la moyenne de la série privée de cette valeur . On dit qu'il s'agit d'une **moyenne élaguée**.

C) MEDIANE *(pour une série ordonnée)*

La médiane est une valeur Me du caractère qui partage la population en deux sous-ensembles de même effectif . Les éléments du premier sous-ensemble correspondent à des valeurs du caractère inférieures ou égales à Me, ceux du second correspondent à des valeurs du caractère supérieures ou égales à Me .

- Si l'effectif total *n* est impair, la médiane est la valeur du caractère située au rang $\frac{n+1}{2}$

- Si l'effectif total *n* est pair, la médiane est tout nombre situé entre la valeur du caractère occupant le rang $\frac{n}{2}$ et la valeur du caractère occupant le rang $\frac{n}{2} + 1$

Dans le cas où les observations sont regroupées en classe, la médiane est la valeur (théorique) du caractère occupant le rang $\frac{n}{2}$ que l'on détermine grâce au polygône des effectifs cumulés croissants.

Ex :

- Ordonnons de la plus basse à la plus haute les 9 notes les plus basses obtenues par les élèves :

  4 4 6 6 7 8 8 9 9

  La note médiane est 7, c'est la 5 ème note de la série ordonnée.

- Ordonnons maintenant l'ensemble des notes obtenues par les élèves :

  4 4 6 6 7 8 8 9 9 9 9 10 10 10 11 11 12 12 12 12 13 13 13 14 14 16 16 17

  On peut prendre pour note médiane tout nombre compris strictement entre 10 et 11. On prend par exemple Me = 10,5 .

4) QUELQUES PROPRIETES DE LA MOYENNE

Comme la moyenne au devoir de la classe a été relativement faible, le professeur, dans sa grande générosité (!), propose trois possibilités pour augmenter cette moyenne :

- Il décide d'ajouter 0,5 points à chaque devoir :
  Que devient la moyenne ? Elle augmente de 0,5 points.

Soit k un réel quelconque.
La moyenne des nombres $x_1 + k, x_2 + k, \dots, x_n + k$ est égale à $\overline{x} + k$

Preuve :

$$\frac{(x_1 + k) + (x_2 + k) + \dots + (x_n + k)}{n} = \frac{x_1 + x_2 + \dots + x_n}{n} + n\frac{k}{n} = \overline{x} + k$$

- Il décide de multiplier chaque note par 1,1.
  Que devient la moyenne ? Elle est multipliée par 1,1.

Soit λ un réel quelconque.
La moyenne des nombres $\lambda x_1, \lambda x_2, \dots, \lambda x_n$ est égale à $\lambda \overline{x}$

Preuve :

$$\frac{\lambda x_1 + \lambda x_2 + \dots + \lambda x_n}{n} = \lambda \frac{x_1 + x_2 + \dots + x_n}{n} = \lambda \overline{x}$$

- Il décide de faire une interrogation notée sur 10 dont voici les résultats des élèves classés dans le même ordre que pour le premier devoir :
  6 10 8 8,5 8 7,5 6 7,5 8 9,5 6,5 7,5 6,5 10 9 9 7,5 9 9 10 10 6 9,5 9 5 10 2 10
  - Quelle est la moyenne sur 10 de cette interrogation ?
  $\overline{y} \approx 8$.
  - Ajouter chaque note de l'interrogation à la note du devoir commun pour obtenir une troisième série de notes, mais sur un total de 30. Déterminer la moyenne de cette nouvelle série.
  La moyenne sur 30 de ces notes est environ 18,5 , c'est à dire $\overline{x} + \overline{y}$.

Soit $\overline{y}$ la moyenne des nombres $y_1, y_2, \dots, y_n$
La moyenne des nombres $x_1 + y_1, x_2 + y_2, \dots, x_n + y_n$ est égale à $\overline{x} + \overline{y}$

Preuve :

$$\frac{(x_1 + y_1) + (x_2 + y_2) + \dots + (x_n + y_n)}{n} = \frac{x_1 + x_2 + \dots + x_n}{n} + \frac{y_1 + y_2 + \dots + y_n}{n}$$
$$= \overline{x} + \overline{y}$$

- Pour le même devoir commun, la classe de 2nde3 contenant 27 élèves a eu une moyenne de 11,3.
  Quelle est la moyenne à ce devoir des élèves des deux classes ?
  $$\frac{28 \times 10{,}5 + 27 \times 11{,}3}{28 + 27} \approx 10{,}9$$

On répartit n nombres en *k* **sous-groupes disjoints**. Le premier groupe contient n_1 éléments de moyenne $\overline{x}_1$, le deuxième n_2 éléments de moyenne $\overline{x}_2$, …, le dernier n_k éléments de moyenne $\overline{x}_k$. Alors la moyenne $\overline{x}$ des n nombres peut-être calculée par :

$$\overline{x} = \frac{\overline{x}_1 n_1 + \overline{x}_2 n_2 + \dots + \overline{x}_k n_k}{n}$$

5) Une caracteristique de dispersion : L'ETENDUE

… indique de quelle façon les valeurs du caractère sont groupées (plus ou moins resserrées) autour des valeurs centrales . L'étendue est la plus simple de ces caractéristiques.

L'étendue e mesure l'écart entre la plus grande et la plus petite valeur prise par la variable.

Ex :

- (tab1) e =

- (tab3) e =